

DTIC FILE COPY

41

SECURITY CLASSIFICATION OF THIS PAGE

## REPORT DOCUMENTATION PAGE

AD-A201 715

		1b RESTRICTIVE MARKINGS	
		3. DISTRIBUTION/AVAILABILITY OF REPORT	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) LIDS-P-1825		5. MONITORING ORGANIZATION REPORT NUMBER(S)	
6a. NAME OF PERFORMING ORGANIZATION M.I.T.	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Office of Naval Research	
6c. ADDRESS (City, State, and ZIP Code) Massachusetts Institute of Technology Cambridge, MA 02139		7b. ADDRESS (City, State, and ZIP Code) 800 North Quincy St. Arlington, VA 22217-5000	
8a. NAME OF FUNDING/SPONSORING ORGANIZATION	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-84-K-0519	
8c. ADDRESS (City, State, and ZIP Code)		10. SOURCE OF FUNDING NUMBERS	
		PROGRAM ELEMENT NO.	PROJECT NO.
		TASK NO.	WORK UNIT ACCESSION NO.
		NR 649-003	
11. TITLE (Include Security Classification) EXPLICIT SOLUTIONS FOR SOME SIMPLE DECENTRALIZED DETECTION PROBLEMS			
12. PERSONAL AUTHOR(S) G. Polychronopoulos, J.N. Tsitsiklis			
13a. TYPE OF REPORT Technical	13b. TIME COVERED FROM TO	14. DATE OF REPORT (Year, Month, Day) October 1988	15. PAGE COUNT 20
16. SUPPLEMENTARY NOTATION			
17. COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	SUB-GROUP	
19. ABSTRACT (Continue on reverse if necessary and identify by block number) We consider a decentralized detection problem in which a number of identical sensors transmit a finite-valued function of their observations to a fusion center which makes a final decision on one of M alternative hypotheses. We consider the case where the number of sensors is large and we derive (asymptotically) optimal rules for determining the messages of the sensors, for the case where the observations are generated from a simple and symmetrical set of discrete distributions. We also consider the tradeoff between the number of sensors and the communication rate of each sensor where there is a constraint on the total communication rate from the sensors to the fusion center.			
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT <input type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS PRT <input type="checkbox"/> DTIC USERS		21. ABSTRACT SECURITY CLASSIFICATION E	
22a. NAME OF RESPONSIBLE INDIVIDUAL		22b. TELEPHONE (Include Area Code) 22c. OFFICE SYMBOL	

DTIC  
ELECTE  
OCT 28 1988

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted  
All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE

U.S. Government Printing Office: 1985-507-047

88 10 20 002

# EXPLICIT SOLUTIONS FOR SOME SIMPLE DECENTRALIZED DETECTION PROBLEMS<sup>1</sup>

George Polychronopoulos<sup>2</sup>

John N. Tsitsiklis<sup>2</sup>

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

## Abstract

We consider a decentralized detection problem in which a number of identical sensors transmit a finite-valued function of their observations to a fusion center which makes a final decision on one of  $M$  alternative hypotheses. We consider the case where the number of sensors is large and we derive (asymptotically) optimal rules for determining the messages of the sensors, for the case where the observations are generated from a simple and symmetrical set of discrete distributions. We also consider the tradeoff between the number of sensors and the communication rate of each sensor when there is a constraint on the total communication rate from the sensors to the fusion center.

1. Research supported by the M.I.T.-Lincoln Laboratory and by the ONR under Contract N00014-84-K-0519 (NR 649-003).

2. Laboratory for Information and Decision Systems and the Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

## 1. PROBLEM DEFINITION

The decentralized detection problem is defined as follows. There are  $M \geq 2$  hypotheses  $H_1, H_2, \dots, H_M$  with known a priori probabilities  $P(H_j) > 0$  and  $N$  sensors. Each sensor  $i$  obtains an observation  $y_i$ , where  $y_i$  is a random variable taking values in a set  $Y$ . We assume that  $y_1, \dots, y_N$  are conditionally independent (given the true hypothesis) and identically distributed with known conditional distributions  $P_Y(\cdot | H_j)$ . Each sensor  $i$  evaluates a  $D$ -valued message  $u_i \in \{1, \dots, D\}$ , as a function of its observation, and transmits it to a fusion center. Finally, the fusion center declares one of the alternative hypotheses to be true (Fig. 1).

Let  $\gamma_i : Y \mapsto \{1, \dots, D\}$ ,  $i = 1, 2, \dots, N$ , be the function (to be called a *decision rule*) used by the  $i$ th sensor to determine its message  $u_i$ ; that is,  $u_i = \gamma_i(y_i)$ . Let  $u_0 \in \{1, \dots, M\}$  be the decision of the fusion center. This decision is made according to a decision rule  $\gamma_0 : \{1, \dots, D\}^N \mapsto \{1, \dots, M\}$ ; that is,  $u_0 = \gamma_0(u_1, \dots, u_N)$ . We say that the fusion center makes an error if  $u_0 = i$  and  $H_i$  is not the true hypothesis. The probability of error is completely determined by the statistics of the observations and by the decision rules  $\gamma_0, \gamma_1, \dots, \gamma_N$ ; it will be denoted by  $J_N(\gamma_0, \dots, \gamma_N)$ . Our problem is to choose the decision rules  $\gamma_0, \gamma_1, \dots, \gamma_N$  of the sensors and of the fusion center so as to minimize the probability of error.

The above described problem and its variations have attracted substantial interest [TeS81], [KuP82], [EkT82], [Tsi84], [TeV84], [TsA85], [PaA86], [HoV86], [ChV86], [Sad86], [Sri86a], [Sri86b], [ReN87a], [ReN87b], [TVB87]. It was first introduced in [TeS81] for the case of two hypotheses ( $M = 2$ ), two sensors ( $N = 2$ ), binary messages ( $D = 2$ ), and for a fixed choice of the fusion center's decision rule  $\gamma_0$ . It was shown in [TeS81] that under the conditional independence assumption, each sensor should evaluate its message  $u_i$  using a likelihood ratio test with an appropriate threshold. (This conclusion is not valid if the conditional independence assumption is removed in which case the problem becomes computationally intractable [TsA85].) The optimal thresholds in the likelihood ratio tests of the different sensors can be obtained by solving a system of nonlinear equations. It is important to emphasize that the optimal decision rules for the decentralized problem are not, in general, the same as those that would be derived using the classical theory, independently for each sensor. This is because the optimal decision rules are chosen so as to optimize systemwide

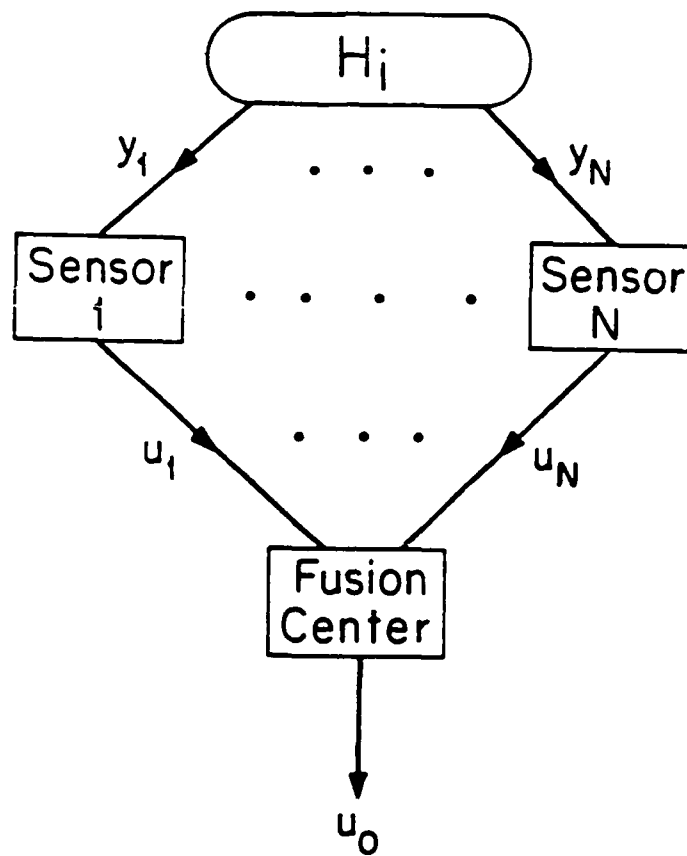


Figure 1

performance, as opposed to the performance of each individual sensor.

The performance of a decentralized detection system is generally inferior to that of a centralized system in which all raw data available are transmitted to the fusion center, due to the loss of information in the local processing. However, decentralized detection is often more practical due to the reduction of the communication requirements, as well as because the processing of the data is shared by a number of different processors. On the other hand, decentralized detection problems are qualitatively different and much more difficult than the corresponding centralized detection problems. For this reason, there are very few such problems that have been solved analytically. In fact, most of the theoretical research available is limited to the derivation of necessary conditions for optimality, and these can only be solved numerically. In contrast, in this paper, we identify a special case for which an explicit solution can be obtained analytically.

We now define the particular problem to be studied. We assume that there is a one-to-one correspondence between observations and hypotheses and, more specifically,  $Y = \{1, \dots, M\}$ . We assume that the conditional distribution of the observation  $y$  of any sensor is given by

$$\Pr(y = i | H_j) = P_Y(i | H_j) = \begin{cases} \epsilon, & \text{if } j \neq i, \\ 1 - (M - 1)\epsilon, & \text{if } j = i, \end{cases}$$

where  $\epsilon$  is a scalar satisfying  $0 < \epsilon \leq 1/(M - 1)$ . In other words, the observation of a sensor indicates the true hypothesis with probability  $1 - (M - 1)\epsilon$ , or it indicates a false hypothesis in which case each one of the false hypotheses is equally likely (probability  $\epsilon$ ). Furthermore, we assume that the number of sensors is large and we will be looking for an asymptotic solution, as  $N \rightarrow \infty$ .

Our model is undoubtedly too structured to be an exact representation of a realistic problem, the main drawback being the assumption that there is a one-to-one correspondence between hypotheses and possible observations. This assumption becomes fairly reasonable, however, in the following situation (see Fig. 2). Each sensor  $i$  receives some observations  $z_i$  that it processes in some predetermined way, and comes up with a preliminary decision  $y_i \in \{1, \dots, M\}$  on the identity of the true hypothesis. Then, each sensor  $i$  transmits to the fusion center a function  $\gamma_i(y_i)$  of its preliminary decision  $y_i$ . Notice that we are restricting here the message to be a function of the processed observations instead of the raw observations. While such a restriction may result to some loss of performance, it is quite natural in certain contexts, especially if each sensor has a reason to

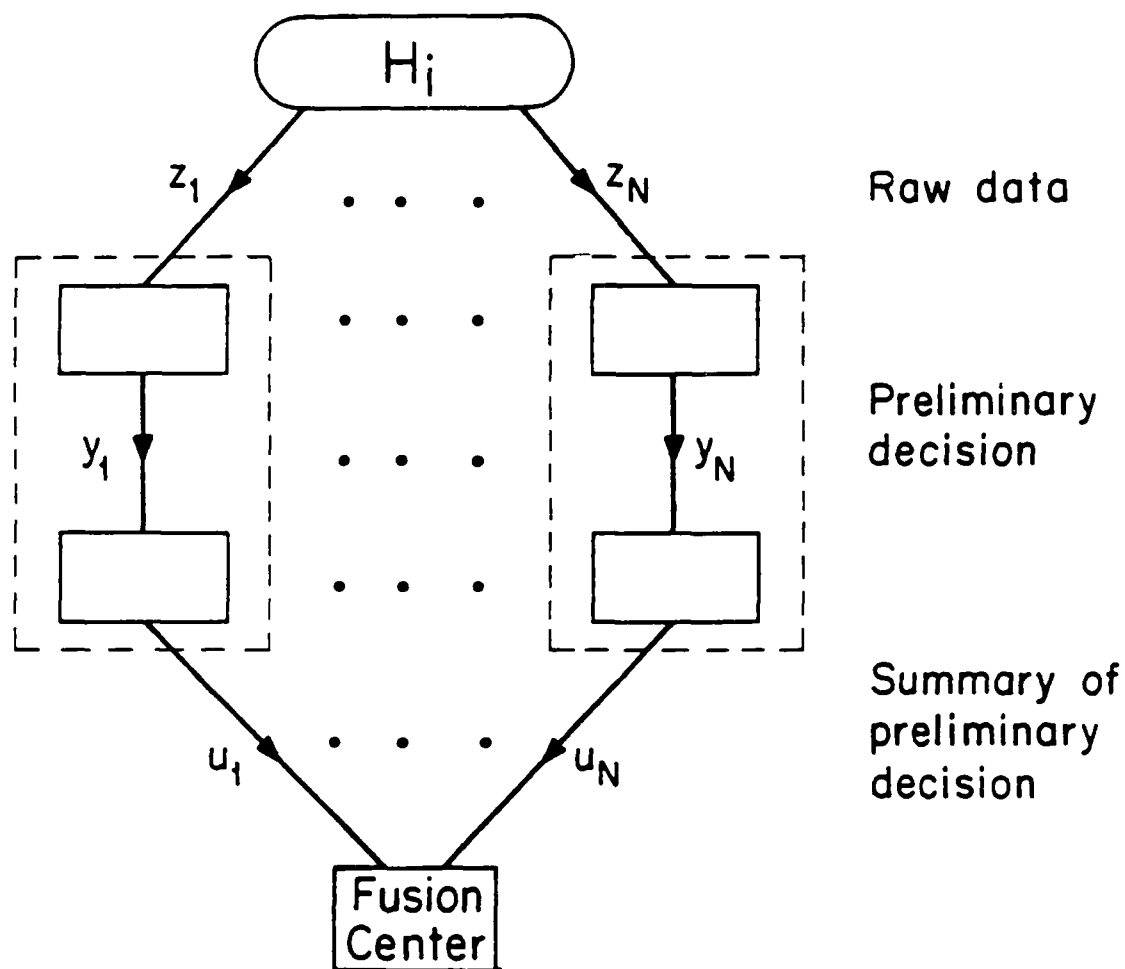


Figure 2

come up with a preliminary decision in a timely manner.

The above discussion notwithstanding, our interest in this particular problem arises mainly from the fact that an explicit solution can be obtained, as will be demonstrated in the sequel. Furthermore, the solution to be derived provides insights and intuition on the nature of optimal solutions to more general problems for which explicit solutions are not possible. Such insights are very valuable because they can suggest interesting numerical experiments and heuristic guidelines for coping with more difficult problems.

The remainder of this paper is organized as follows. In Section 2, we outline some results from [Tsi88] that will be needed later. In Section 3, we introduce some notation and terminology, and some simple preliminary facts. In Section 4, a complete solution is derived for the case where the noise parameter  $\epsilon$  is small and the number of sensors is large. In Section 5, we provide a partial extension of the results of Section 4 to the case of a general noise parameter  $\epsilon$ . Finally, in Section 6, we study the tradeoff between the number of sensors and the communication rate of each sensor when there is a constraint on the total communication rate from the sensors to the fusion center.

## 2. BACKGROUND.

As mentioned in the introduction, we will be looking for an asymptotic solution to our problem, as the number of sensors  $N$  becomes very large. The basic theory concerning such an asymptotic solution has been developed in [Tsi88] and we review here the facts that will be needed. Some experimentation [Pol88] has shown that the asymptotically optimal decision rules perform reasonably well for moderate numbers of sensors.

We use  $\Gamma$  to denote the set of all possible decision rules. Due to the finiteness of the observation set  $Y$  and of the message set  $\{1, \dots, D\}$ , it is seen that the set  $\Gamma$  is also finite. We introduce the shorthand notation  $\gamma^N$  to denote a possible choice  $(\gamma_0, \gamma_1, \dots, \gamma_N)$  of decision rules for the  $N$ -sensor problem. With a reasonable choice of  $\gamma^N$ , the probability of error  $J_N(\gamma^N)$  converges exponentially to zero as  $N$  increases. For this reason, we focus on the exponent of the error probability, defined by

$$r_N(\gamma^N) = \frac{\log J_N(\gamma^N)}{N}. \quad (1)$$

Let  $R_N = \inf_{\gamma^N} r_N(\gamma^N)$ , where the infimum is taken over all possible choices of decision rules for the  $N$ -sensor problem. Thus,  $R_N$  is the optimal exponent. As  $N$  tends to infinity,  $R_N$  has a limit [Tsi88] which will be denoted by  $\Lambda^*$ . In the sequel, we will be concerned with choosing the decision rules so that the corresponding error exponent approaches the optimal exponent  $\Lambda^*$ .

Consider a sensor that uses a particular decision rule  $\gamma \in \Gamma$ . Conditioned on  $H_i$ , the probability that the transmitted message takes a particular value  $d \in \{1, \dots, D\}$  is given by  $\Pr(\gamma(y) = d \mid H_i)$ . For every  $i, j \in \{1, \dots, M\}$  and every decision rule  $\gamma \in \Gamma$ , we define a function  $\mu_{ij}(\gamma, \cdot) : [0, 1] \mapsto [-\infty, +\infty)$  by

$$\mu_{ij}(\gamma, s) = \log \left[ \sum_{d=1}^D \left( \Pr(\gamma(y) = d \mid H_i) \right)^{1-s} \left( \Pr(\gamma(y) = d \mid H_j) \right)^s \right]. \quad (2)$$

(The convention  $0^0 = 0$  is used in this formula.) It is easily verified that  $\mu_{ij}(\gamma, s) \leq 0$  for every  $i, j$ ,  $\gamma \in \Gamma$ ,  $s \in [0, 1]$ , and it is also known that  $\mu_{ij}(\gamma, s)$  is a convex function of  $s$ , for every  $i, j$ ,  $\gamma \in \Gamma$  [SGB67]. Furthermore, as long as there exists some  $y \in Y$  such that  $P_Y(y \mid H_i) \cdot P_Y(y \mid H_j) \neq 0$ , then  $\mu_{ij}(\gamma, s) > -\infty$ , for every  $s \in [0, 1]$ . This turns out to be always the case for our problem except for the uninteresting situation where  $M = 2$  and  $\epsilon = 1$ .

The optimal exponent is given by [Tsi88]

$$\Lambda^* = \min_{\{x_\gamma \mid \gamma \in \Gamma\}} \max_{\{(i,j) \mid i \neq j\}} \min_{s \in [0,1]} \sum_{\gamma \in \Gamma} x_\gamma \mu_{ij}(\gamma, s), \quad (3)$$

where the outer minimization is carried out over all choices of  $\{x_\gamma \mid \gamma \in \Gamma\}$  satisfying  $x_\gamma \geq 0$  for all  $\gamma \in \Gamma$ , and  $\sum_{\gamma \in \Gamma} x_\gamma = 1$ . In the sequel, we use  $x$  to denote a vector  $\{x_\gamma \mid \gamma \in \Gamma\}$ . Furthermore, we use  $X$  to denote the set of all such vectors which satisfy the constraints just stated.

The variable  $x_\gamma$  in Eq. (3) should be interpreted as the fraction of the sensors that use decision rule  $\gamma$ . More specifically, let us fix some  $x \in X$ . For each  $\gamma \in \Gamma$ , let  $\lfloor Nx_\gamma \rfloor$  sensors use decision rule  $\gamma$ . (If for some  $\gamma$  the value of  $Nx_\gamma$  is not integer this determines the decision rules for fewer than  $N$  sensors. However, the remaining sensors constitute a vanishingly small fraction of the total, as  $N \rightarrow \infty$ , and are inconsequential.) Then, the asymptotic exponent (as  $N \rightarrow \infty$ ) of the probability of error is given by [Tsi88]

$$\max_{\{(i,j) \mid i \neq j\}} \min_{s \in [0,1]} \sum_{\gamma \in \Gamma} x_\gamma \mu_{ij}(\gamma, s). \quad (4)$$



In particular, if the fractions  $x_\gamma$  are chosen to minimize the exponent in Eq. (4), then the optimal exponent  $\Lambda^*$  is obtained [compare with Eq. (3)]. Notice that the problem formulation has taken a somewhat different, but equivalent, form: instead of choosing the decision rule of each sensor, we are now trying to choose the fraction  $x_\gamma$  of the sensors that use a given decision rule  $\gamma \in \Gamma$ .

Equation (4) has a simple interpretation. The quantity  $\min_{s \in [0,1]} \sum_{\gamma \in \Gamma} x_\gamma \mu_{ij}(\gamma, s)$  is the exponent in the Chernoff bound for the probability of confusing hypotheses  $H_i$  and  $H_j$  ([VaT68], [SGB67]), and such a bound is known to be asymptotically tight. The maximization over all  $i$  and  $j$  in Eq. (4) corresponds to the fact that the dominant term in the probability of error comes from the worst (i.e., the largest) of the exponents corresponding to the different pairs.

The outer minimization in Eq. (3) appears to be simple because it involves linear constraints and a cost function which is linear in the variables  $x_\gamma$ . However, the inner minimization (with respect to  $s$ ) severely complicates the computation of  $\Lambda^*$  and of the optimal values of the variables  $x_\gamma$ . In the next two sections, we get around this difficulty by exploiting the symmetry of the problem to remove the dependence on  $s$ .

### 3. PRELIMINARIES.

Consider a decision rule  $\gamma : Y \mapsto \{1, \dots, D\}$  and let  $Y_{d,\gamma} = \{y \mid \gamma(y) = d\}$ . We notice that the sets  $Y_{d,\gamma}$ ,  $d = 1, \dots, D$ , are disjoint and their union equals  $Y$ . Thus, a decision rule determines a partition of  $Y$  into  $D$  disjoint sets. It is possible that two different functions  $\gamma : Y \mapsto \{1, \dots, D\}$  and  $\gamma' : Y \mapsto \{1, \dots, D\}$  determine the same partition. [For example, consider the case where  $\gamma'(y) = D + 1 - \gamma(y)$ .] On the other hand, if  $\gamma$  and  $\gamma'$  determine the same partition, then each one of the messages  $\gamma(y)$  and  $\gamma'(y)$  conveys the same information to the fusion center, and the two decision rules can be considered equivalent. From now on, we will not distinguish between equivalent decision rules and we will consider them to be identical. We are therefore adopting the alternative definition that a decision rule is a partition of  $Y$  into subsets  $Y_{1,\gamma}, \dots, Y_{D,\gamma}$ . We assume that the sets  $Y_{d,\gamma}$  are arranged in order of increasing cardinality; that is,  $|Y_{1,\gamma}| \leq \dots \leq |Y_{D,\gamma}|$ .

**Definition:** Two observations  $i, j \in Y$  are *separated* by a decision rule  $\gamma$  if  $i$  and  $j$  belong to different elements  $Y_{d,\gamma}$  of the partition corresponding to  $\gamma$ . We let  $\Gamma_{i,j}$  be the set of all  $\gamma \in \Gamma$  that

separate  $i$  and  $j$ . The number of separations corresponding to a decision rule  $\gamma$  is defined as the number of (unordered) pairs of observations  $i, j \in Y$  which are separated by  $\gamma$ .

Notice that an  $M$ -ary hypothesis testing problem can be viewed as a collection of several binary hypothesis testing problems, one for each pair of hypotheses. The number of separations corresponding to a decision rule  $\gamma$  can be interpreted as the number of binary problems for which a message  $\gamma(y_i)$  provides useful information.

**Definition:** Let  $\delta_1, \dots, \delta_D$  be a collection of nonnegative integers satisfying  $\delta_1 \leq \delta_2 \leq \dots \leq \delta_D$  and  $\sum_{d=1}^D \delta_d = M$ . The class  $C^{\delta_1, \dots, \delta_D}$  is the set of all  $\gamma \in \Gamma$  such that  $|Y_{d,\gamma}| = \delta_d$  for every  $d$ .

These definitions are illustrated in Fig. 3.

Let  $L$  be the number of different classes. In order to facilitate notation, we assume that the different classes have been arranged according to some arbitrary order and we will use the simpler notation  $C_\ell$  to denote the  $\ell$ th class,  $\ell = 1, \dots, L$ . Thus, the set  $\Gamma$  of all decision rules is equal to  $\cup_{\ell=1}^L C_\ell$ .

It is seen that the number of separations is the same for all decision rules belonging to the same class  $C_\ell$  [see Fig. 3], and will be denoted by  $S_\ell$ . In particular,

$$S_\ell = \frac{1}{2} \sum_{d=1}^D \delta_d (M - \delta_d), \quad (5)$$

where  $\delta_1, \dots, \delta_D$  are such that  $C_\ell = C^{\delta_1, \dots, \delta_D}$ . [The factor 1/2 in Eq. (5) is present because otherwise each unordered pair would be counted twice.]

Let  $Q_\ell$  be the cardinality of the set of all triples  $(i, j, \gamma)$  such that  $\gamma \in C_\ell$  and  $\gamma$  separates  $i$  and  $j$ . [The two triples  $(i, j, \gamma)$  and  $(j, i, \gamma)$  are only counted once.] Since the number of separations corresponding to any  $\gamma \in C_\ell$  is  $S_\ell$ , we see that  $Q_\ell = |C_\ell| \cdot S_\ell$ . On the other hand, every pair  $(i, j)$  is separated by exactly  $|C_\ell \cap \Gamma_{ij}|$  elements of  $C_\ell$ . By symmetry, the cardinality of  $C_\ell \cap \Gamma_{ij}$  is the same for every  $i$  and  $j$ . Furthermore, since there exist  $M(M-1)/2$  different (unordered) pairs  $(i, j)$ , we conclude that  $Q_\ell = |C_\ell \cap \Gamma_{ij}| \cdot M(M-1)/2$ . By equating the two alternative expressions for  $Q_\ell$ , we obtain

$$\frac{|C_\ell \cap \Gamma_{ij}|}{|C_\ell|} = \frac{2S_\ell}{M(M-1)}, \quad (6)$$

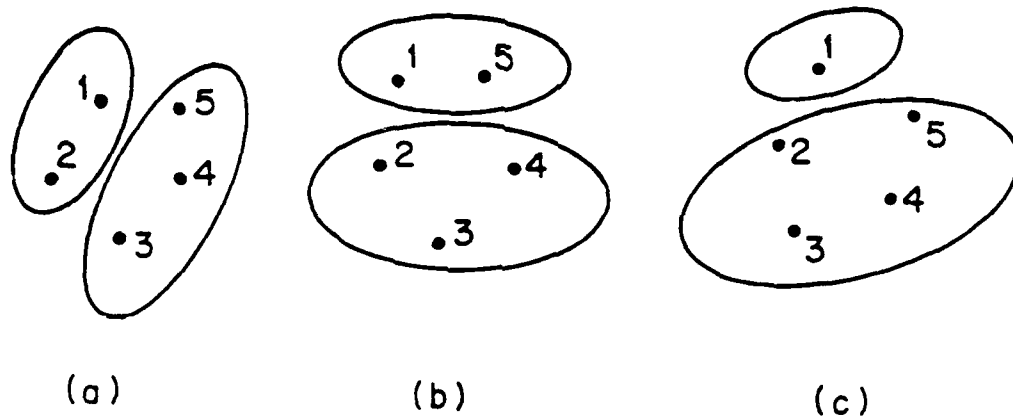


Figure 3

Let  $M = 5$ ,  $D = 2$ . The decision rules in (a) and (b) belong to the class  $C^{2,3}$  and the corresponding number of separations is 6. The decision rule in (c) belongs to the class  $C^{1,4}$  and the corresponding number of separations is 4.

a fact that will be useful later.

We now derive the form of the functions  $\mu_{ij}(\gamma, s)$ . Suppose that  $i \in Y_{\eta, \gamma}$  and  $j \in Y_{\zeta, \gamma}$ . Using the notation  $\delta_\eta = |Y_{\eta, \gamma}|$  and  $\delta_\zeta = |Y_{\zeta, \gamma}|$ , it is seen [cf. Eq. (2)] that

$$\mu_{ij}(\gamma, s) = \log \left[ (1 - (M - \delta_\eta)\epsilon)^{1-s} (\delta_\eta \epsilon)^s + (\delta_\zeta \epsilon)^{1-s} (1 - (M - \delta_\zeta)\epsilon)^s + (M - \delta_\eta - \delta_\zeta)\epsilon \right], \quad (7)$$

if  $\eta \neq \zeta$ ,

and

$$\mu_{ij}(\gamma, s) = 0, \quad \text{if } \eta = \zeta. \quad (8)$$

Notice that the case  $\eta = \zeta$  [cf. Eq. (8)] corresponds to the case  $\gamma \notin \Gamma_{ij}$ . Finally, from either Eq. (2) or Eq. (7), it is seen that

$$\mu_{ij}(\gamma, s) = \mu_{ji}(\gamma, 1 - s), \quad (9)$$

which will be useful later.

#### 4. THE SMALL NOISE CASE.

In this section, we derive the solution of the problem under consideration for the case where the noise parameter  $\epsilon$  is small. This is accomplished by showing that the minimum with respect to  $s$  in Eq. (3) is approximately attained for  $s = 1/2$ , which allows us to eliminate  $s$ .

**Lemma 1:** Fix some  $\epsilon_0$  such that  $0 < \epsilon_0 < 1/(M - 1)$ . Then, there exist constants  $G_1$  and  $G_2$  such that, for every  $\epsilon \in (0, \epsilon_0)$ , every  $i, j \in \{1, \dots, M\}$  such that  $i \neq j$ , and every  $x \in X$ , we have

$$G_1 + \frac{1}{2} \log \epsilon \sum_{\gamma \in \Gamma_{ij}} x_\gamma \leq \min_{s \in [0, 1]} \sum_{\gamma \in \Gamma} x_\gamma \mu_{ij}(\gamma, s, \epsilon) \leq G_2 + \frac{1}{2} \log \epsilon \sum_{\gamma \in \Gamma_{ij}} x_\gamma.$$

**Proof:** We first prove the right-hand side inequality. Consider some  $\gamma \in \Gamma_{ij}$  and suppose that  $i \in Y_{\eta, \gamma}$ ,  $j \in Y_{\zeta, \gamma}$ . We have [cf. Eq. (7)]

$$\begin{aligned} e^{\mu_{ij}(\gamma, 1/2)} &= (1 - (M - \delta_\eta)\epsilon)^{1/2} (\delta_\eta \epsilon)^{1/2} + (\delta_\zeta \epsilon)^{1/2} (1 - (M - \delta_\zeta)\epsilon)^{1/2} + (M - \delta_\eta - \delta_\zeta)\epsilon \\ &\leq (\delta_\eta \epsilon)^{1/2} + (\delta_\zeta \epsilon)^{1/2} + (M - \delta_\eta - \delta_\zeta)\epsilon^{1/2} \leq H_2 \epsilon^{1/2}, \end{aligned}$$

where  $H_2 = \delta_\eta^{1/2} + \delta_\zeta^{1/2} + M - \delta_\eta - \delta_\zeta > 1$ . Taking logarithms, we obtain  $\mu_{ij}(\gamma, 1/2) \leq G_2 + (\log \epsilon)/2$ , where  $G_2 = \log H_2 > 0$ . Furthermore, if  $\gamma \notin \Gamma_{ij}$ , we have  $\mu_{ij}(\gamma, 1/2) = 0$  [cf. Eq. (8)]. It follows

that

$$\sum_{\gamma \in \Gamma} x_{\gamma} \mu_{ij}(\gamma, 1/2) \leq G_2 + \frac{1}{2} \log \epsilon \sum_{\gamma \in \Gamma_{ij}} x_{\gamma}.$$

If a minimization over  $s$  is carried out, the resulting value is no larger than the one corresponding to  $s = 1/2$ , and this proves the right-hand side inequality.

We now prove the left-hand side inequality. We fix some  $i, j$ , some  $\gamma \in \Gamma_{ij}$ , and some  $s \in [0, 1/2]$ . We assume again that  $i \in Y_{\eta, \gamma}$  and  $j \in Y_{\epsilon, \gamma}$ . We have

$$\begin{aligned} e^{\mu_{ij}(\gamma, s)} &= (1 - (M - \delta_{\eta})\epsilon)^{1-s} (\delta_{\eta}\epsilon)^s + (\delta_{\epsilon}\epsilon)^{1-s} (1 - (M - \delta_{\epsilon})\epsilon)^s + (M - \delta_{\eta} - \delta_{\epsilon})\epsilon \\ &\geq (1 - (M - \delta_{\eta})\epsilon)^{1-s} (\delta_{\eta}\epsilon)^s \geq (1 - (M - \delta_{\eta})\epsilon)\epsilon^s \geq (1 - (M - 1)\epsilon_0)\epsilon^s = H_1\epsilon^s \geq H_1\epsilon^{1/2}, \end{aligned}$$

where  $H_1 = 1 - (M - 1)\epsilon_0 > 0$ . Taking logarithms, we obtain  $\mu_{ij}(\gamma, s) \geq G_1 + (\log \epsilon)/2$ , where  $G_1 = \log H_1 < 0$ . The same conclusion is obtained by a symmetrical argument for the case  $s \in [1/2, 1]$ . Using again the fact that  $\mu_{ij}(\gamma, s) = 0$  if  $\gamma \notin \Gamma_{ij}$ , we obtain

$$\min_{s \in [0, 1]} \sum_{\gamma \in \Gamma} x_{\gamma} \mu_{ij}(\gamma, s) \geq \sum_{\gamma \in \Gamma} x_{\gamma} \min_{s \in [0, 1]} \mu_{ij}(\gamma, s) \geq \sum_{\gamma \in \Gamma_{ij}} x_{\gamma} \left( \frac{1}{2} \log \epsilon + G_1 \right) \geq G_1 + \frac{1}{2} \log \epsilon \sum_{\gamma \in \Gamma_{ij}} x_{\gamma},$$

which completes the proof. **Q.E.D.**

We notice that as  $\epsilon$  approaches zero,  $\log \epsilon$  tends to  $-\infty$ , while the constants  $G_1, G_2$  of Lemma 1 remain unchanged. Therefore, by retaining the dominant term,  $\Lambda^*$  can be approximated, for small  $\epsilon$ , by

$$\Lambda^* = \frac{1}{2} \min_{z \in X} \max_{\{(i, j) | i \neq j\}} \log \epsilon \sum_{\gamma \in \Gamma_{ij}} x_{\gamma}. \quad (10)$$

Since  $\log \epsilon$  is negative, an equivalent optimization problem is

$$\max_{z \in X} \min_{\{(i, j) | i \neq j\}} \sum_{\gamma \in \Gamma_{ij}} x_{\gamma}. \quad (11)$$

We now derive the solution of (11).

**Proposition 1:** Let  $S^* = \max_{\ell} S_{\ell}$ . Then, a vector  $z \in X$  is an optimal solution of the problem (11) if and only if the following two conditions hold:

- (i) The value of  $\sum_{\gamma \in \Gamma_{ij}} x_{\gamma}$  is the same for every pair  $(i, j)$  such that  $i \neq j$ .
- (ii) If  $\gamma \in C_{\ell}$  and  $S_{\ell} < S^*$ , then  $x_{\gamma} = 0$ .

Furthermore, the optimal value of (11) is equal to  $2S^*/(M(M-1))$ .

**Proof:** Suppose that a vector  $x^* \in X$  satisfies conditions (i) and (ii), and let  $c$  be such that  $c = \sum_{\gamma \in \Gamma_{ij}} x_\gamma^*$ , for  $i \neq j$ . Summing over all unordered pairs  $(i, j)$ , we obtain

$$c \frac{M(M-1)}{2} = \sum_{\{(i,j)|i \neq j\}} \sum_{\gamma \in \Gamma_{ij}} x_\gamma^* = \sum_{\gamma \in \Gamma} \sum_{\{(i,j)|\gamma \in \Gamma_{ij}\}} x_\gamma^* = \sum_{\gamma \in \Gamma} S^* x_\gamma^* = S^*.$$

[Here we used the fact that if  $\gamma \in C_\ell$ , then the cardinality of the set  $\{(i, j) \mid \gamma \in \Gamma_{ij}\}$  is  $S_\ell$ , by definition; we then used property (ii) to replace  $S_\ell$  by  $S^*$ .] We conclude that if conditions (i) and (ii) hold, then  $c = 2S^*/(M(M-1))$ .

In order to show that the vector  $x^*$  is actually optimal, it is sufficient to show that

$$\min_{\{(i,j)|i \neq j\}} \sum_{\gamma \in \Gamma_{ij}} x_\gamma \leq \frac{2S^*}{M(M-1)},$$

for every vector  $x \in X$ . We use the elementary fact that the minimum of a set of numbers is no larger than their average, to obtain

$$\begin{aligned} \frac{M(M-1)}{2} \min_{\{(i,j)|i \neq j\}} \sum_{\gamma \in \Gamma_{ij}} x_\gamma &\leq \sum_{\{(i,j)|i \neq j\}} \sum_{\gamma \in \Gamma_{ij}} x_\gamma = \sum_{\ell} \sum_{\gamma \in C_\ell} \sum_{\{(i,j)|\gamma \in \Gamma_{ij}\}} x_\gamma \\ &= \sum_{\ell} \sum_{\gamma \in C_\ell} S_\ell x_\gamma \leq S^* \sum_{\ell} \sum_{\gamma \in C_\ell} x_\gamma = S^*, \end{aligned} \quad (12)$$

as desired. We conclude that  $x^*$  is optimal.

For the converse, let us suppose that a vector  $x \in X$  is optimal. We have already established that the optimal value of the objective function under consideration is equal to  $2S^*/(M(M-1))$ . Therefore, all inequalities in Eq. (12) must be equalities. Since the first inequality in Eq. (12) is not strict, condition (i) follows. Furthermore, since the second inequality in Eq. (12) is not strict, condition (ii) follows. **Q.E.D.**

Using Prop. 1, one optimal solution for the problem (11) is the following. Choose a class  $C_{\ell^*}$  such that  $S_{\ell^*} = S^* = \max_{\ell} S_\ell$  and let

$$x_\gamma = \begin{cases} 0, & \text{if } \gamma \notin C_{\ell^*}, \\ \frac{1}{|C_{\ell^*}|}, & \text{if } \gamma \in C_{\ell^*}. \end{cases} \quad (13)$$

It is seen that this vector  $x$  is feasible ( $x \in X$ ) and satisfies the optimality conditions of Prop. 1. Let us point out that an optimal solution of the problem (11) is in general not unique. The solution provided by Eq. (13) can be singled out because of its special symmetry properties.

The class  $C_{\ell^*}$ , which is a class of decision rules with a maximal number of separations, should be viewed as a "best" class: according to Prop. 1 only decision rules in such a class should be used. This is very intuitive because each  $\gamma \in C_{\ell}$  provides information to the fusion center which is useful in discriminating  $S_{\ell}$  pairs of hypotheses (by the definition of  $S_{\ell}$ ). The larger the value of  $S_{\ell}$ , the larger the contribution of a decision rule  $\gamma \in S_{\ell}$  in discriminating between the different hypotheses.

We now proceed to determine the best class  $C_{\ell^*}$ . Suppose that  $C_{\ell^*} = C^{\delta_1, \dots, \delta_D}$ , for some integer coefficients  $\delta_1, \dots, \delta_D$  whose sum is equal to  $M$ . Suppose that there exist some  $\eta$  and  $\zeta$  such that  $\delta_{\eta} - \delta_{\zeta} > 1$ . Consider a new class  $C_{\ell'} = C^{\delta'_1, \dots, \delta'_D}$ , where  $\delta'_{\eta} = \delta_{\eta} - 1$ ,  $\delta'_{\zeta} = \delta_{\zeta} + 1$ , and  $\delta'_d = \delta_d$  if  $d \neq \eta$  and  $d \neq \zeta$ . Using Eq. (5), we obtain

$$\begin{aligned} 2(S_{\ell'} - S_{\ell^*}) &= \delta'_{\zeta}(M - \delta'_{\zeta}) + \delta'_{\eta}(M - \delta'_{\eta}) - \delta_{\zeta}(M - \delta_{\zeta}) - \delta_{\eta}(M - \delta_{\eta}) \\ &= (\delta_{\zeta} + 1)(M - \delta_{\zeta} - 1) + (\delta_{\eta} - 1)(M - \delta_{\eta} + 1) - \delta_{\zeta}(M - \delta_{\zeta}) - \delta_{\eta}(M - \delta_{\eta}) \\ &= 2(\delta_{\eta} - \delta_{\zeta} - 1) > 0, \end{aligned}$$

which contradicts the optimality of  $S_{\ell^*}$ . This shows that  $|\delta_{\eta} - \delta_{\zeta}| \leq 1$  for all  $\eta, \zeta$ . Given that the average of the coefficients  $\delta_d$  must be equal to  $M/D$ , it follows that for every  $d$  we must have either  $\delta_d = \lfloor M/D \rfloor$  or  $\delta_d = \lceil M/D \rceil$ . In particular, if  $M$  is divisible by  $D$ , then  $\delta_d = M/D$  for every  $d$ . If  $M$  is not divisible by  $D$ , the number of  $\delta_d$ 's for which  $\delta_d = \lfloor M/D \rfloor$  is uniquely determined by the requirement  $\sum_{d=1}^D \delta_d = M$ .

We conclude that with decision rules belonging to the best class  $C_{\ell^*}$ , the corresponding partitions of the observation set  $Y$  are as even as possible. For example, if  $D = 2$  and  $M$  is even, the set  $Y$  is to be partitioned into two subsets with equal cardinalities. Also, for the example of Fig. 3 in which  $M = 5$  and  $D = 2$ , the best class is the class  $C^{2,3}$ . Notice that  $C^{2,3}$  has 10 different elements; thus, an optimal solution is to divide the sensors in ten groups of equal cardinality and letting all the sensors in each group use a particular decision rule belonging to the class  $C^{2,3}$ .

## 5. THE GENERAL CASE.

We now consider the case where  $\epsilon$  does not tend to zero but is fixed instead at some nonzero value in the range  $0 < \epsilon < 1/(M-1)$ . Unfortunately, despite the symmetry of the optimization

problem defining  $\Lambda^*$ , symmetry considerations alone are not sufficient to ascertain that the optimal value of the vector  $x$  possesses symmetry properties similar to the ones obtained in the previous section. We demonstrate this by means of a simple example.<sup>†</sup>

**Example:** Let there be three hypotheses ( $M = 3$ ) and let the messages be binary ( $D = 2$ ). In this case there are exactly three decision rules, the following: the  $i$ th decision rule  $\gamma_i$ ,  $i = 1, 2, 3$ , is defined by  $\gamma_i(i) = 1$  and  $\gamma_i(j) = 2$  if  $j \neq i$ . Notice that  $\mu_{12}(\gamma_3, s) = \mu_{13}(\gamma_2, s) = \mu_{23}(\gamma_1, s) = 0$ , for every  $s$ . Let

$$\nu(s) = \log [(1 - 2\epsilon)^{1-s} \epsilon^s + (2\epsilon)^{1-s} (1 - \epsilon)^s].$$

It is seen [cf. Eq. (2)] that  $\mu_{ij}(\gamma_i, s) = \nu(s)$  and  $\mu_{ij}(\gamma_j, s) = \nu(1 - s)$ , for every  $i \neq j$ . Substituting in Eq. (3), and using the notation  $x_i = x_{\gamma_i}$ , we obtain

$$\begin{aligned} \Lambda^* &= \min_{x \in X} \max \left\{ \min_{s \in [0,1]} [x_1 \mu_{12}(\gamma_1, s) + x_2 \mu_{12}(\gamma_2, s) + x_3 \mu_{12}(\gamma_3, s)], \right. \\ &\quad \min_{s \in [0,1]} [x_1 \mu_{13}(\gamma_1, s) + x_2 \mu_{13}(\gamma_2, s) + x_3 \mu_{13}(\gamma_3, s)], \\ &\quad \left. \min_{s \in [0,1]} [x_1 \mu_{23}(\gamma_1, s) + x_2 \mu_{23}(\gamma_2, s) + x_3 \mu_{23}(\gamma_3, s)] \right\} \\ &= \min_{x \in X} \max \left\{ \min_{s \in [0,1]} [x_1 \nu(s) + x_2 \nu(1 - s)], \right. \\ &\quad \min_{s \in [0,1]} [x_1 \nu(s) + x_3 \nu(1 - s)], \\ &\quad \left. \min_{s \in [0,1]} [x_2 \nu(s) + x_3 \nu(1 - s)] \right\}. \end{aligned}$$

Consider the symmetric solution ( $x_i = 1/3$  for each  $i$ ). The corresponding exponent is seen to be  $\frac{1}{3} \min_{s \in [0,1]} [\nu(s) + \nu(1 - s)] = \frac{2}{3} \nu(\frac{1}{2})$ . (The last equality follows because we are minimizing a convex function which is symmetric around the point  $1/2$ .) Let us now consider the nonsymmetric solution  $x_1 = x_2 = \frac{1}{2}$ ,  $x_3 = 0$ . The corresponding exponent is equal to

$$\max \left\{ \nu\left(\frac{1}{2}\right), \frac{1}{2} \min_{s \in [0,1]} \nu(s) \right\}.$$

In particular, if  $\frac{1}{2} \min_{s \in [0,1]} \nu(s) < \frac{2}{3} \nu(\frac{1}{2})$ , then the symmetric solution is not optimal. We have investigated this issue numerically by computing the value of the exponent corresponding to different vectors  $x \in X$  (over a fairly dense grid of points in  $X$  and for a few different values of  $\epsilon$ ) and

<sup>†</sup> This example also corrects an error in a corresponding example in [Tsi88].



we have reached the conclusion that the symmetric solution is always the optimal one. However, an analytical method for establishing that this is the case is not apparent, even though it can be proved that the symmetric solution is a strict local minimum. (The proof of the latter fact is outlined in the Appendix.)

Without any guaranteed symmetry properties, little progress can be made analytically towards the computation of  $\Lambda^*$ . For this reason, we shall *impose* a symmetry requirement and proceed to solve the problem of Eq. (3) subject to this additional constraint. Motivated by the structure of an optimal solution for the low noise case [cf. Eq. (13)], we require that the value of  $x_\gamma$  be the same for every  $\gamma$  belonging to the same class. Given any vector  $x \in X$  satisfying this requirement, let  $y_\ell = \sum_{\gamma \in C_\ell} x_\gamma$ . We then have  $x_\gamma = y_\ell / |C_\ell|$  for every  $\gamma \in C_\ell$ . Using this expression for  $x_\gamma$ , the minimization problem of Eq. (3) becomes

$$\Lambda^* = \min_{y_1, \dots, y_L} \max_{\{(i,j) | i \neq j\}} \min_{s \in [0,1]} \sum_{\ell=1}^L \frac{y_\ell}{|C_\ell|} \sum_{\gamma \in C_\ell} \mu_{ij}(\gamma, s), \quad (14)$$

where the variables  $y_1, \dots, y_L$  are subject to the constraints  $y_\ell \geq 0$ , for each  $\ell$ , and  $\sum_{\ell=1}^L y_\ell = 1$ .

**Proposition 2:** (a) Fix some class  $C_\ell$ . Then, the value of

$$\frac{1}{|C_\ell \cap \Gamma_{ij}|} \sum_{\gamma \in C_\ell \cap \Gamma_{ij}} \mu_{ij}(\gamma, 1/2)$$

is the same for all  $i, j$  such that  $i \neq j$ , and will be denoted by  $\alpha_\ell$ .

(b) Let  $\ell^*$  be such that  $S_{\ell^*} |\alpha_{\ell^*}| = \max_\ell S_\ell |\alpha_\ell|$ . Then, the choice  $y_{\ell^*} = 1$ , and  $y_\ell = 0$  if  $\ell \neq \ell^*$ , is an optimal solution of the problem (14).

**Proof:** (a) This is evident from the definition of  $\mu_{ij}(\gamma, 1/2)$  and symmetry considerations.

(b) Fix some pair  $(i, j)$ , with  $i \neq j$ . For any  $\gamma \in \Gamma$ , define a new decision rule  $\sigma(\gamma)$  in which the positions of  $i$  and  $j$  in the partition corresponding to  $\gamma$  are interchanged (see Fig. 4). It is seen that  $\sigma$  is a one-to-one and onto mapping of any given class  $C_\ell$  into itself. Furthermore, it follows easily from the definition of  $\mu_{ij}$  that  $\mu_{ij}(\sigma(\gamma), s) = \mu_{ji}(\gamma, s) = \mu_{ij}(\gamma, 1-s)$ . Therefore,

$$\sum_{\gamma \in C_\ell} \mu_{ij}(\gamma, s) = \frac{1}{2} \sum_{\gamma \in C_\ell} [\mu_{ij}(\gamma, s) + \mu_{ij}(\sigma(\gamma), s)] = \frac{1}{2} \sum_{\gamma \in C_\ell} [\mu_{ij}(\gamma, s) + \mu_{ij}(\gamma, 1-s)]. \quad (15)$$

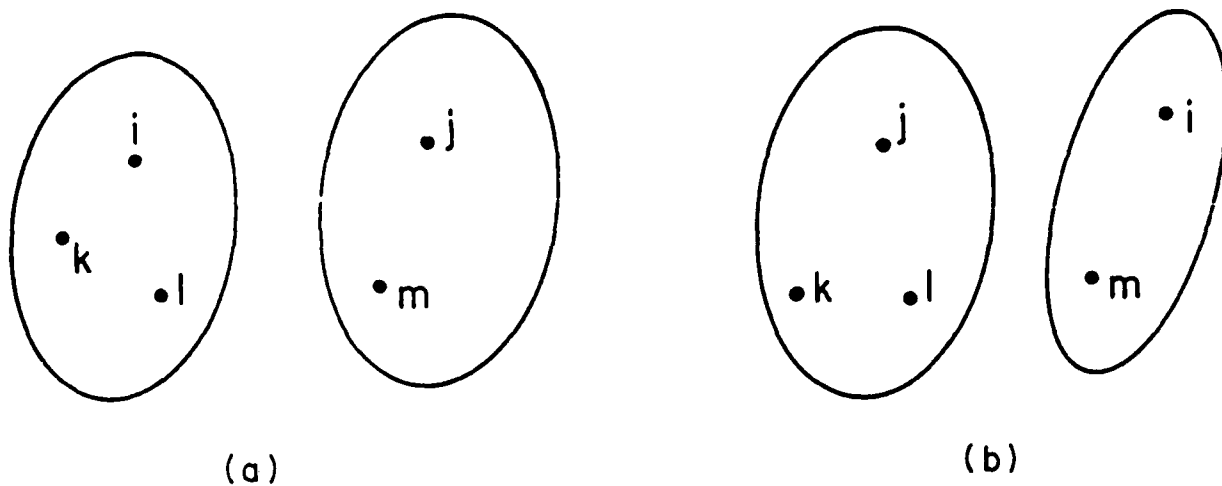


Figure 4

A decision rule  $\gamma$  is shown in (a) and the corresponding decision rule  $\sigma(\gamma)$  (in which the positions of  $i$  and  $j$  are interchanged) is shown in (b).

Thus, the expression in the left-hand side of Eq. (15) is symmetric, as a function of  $s$ , around the value  $s = 1/2$ . It follows that the minimization with respect to  $s$  in Eq. (14) involves a function which is convex and symmetric around the point  $s = 1/2$ . Hence, the minimum is attained at  $s = 1/2$  and Eq. (14) simplifies to

$$\Lambda^* = \min_{y_1, \dots, y_L} \max_{\{(i,j) | i \neq j\}} \sum_{\ell=1}^L \frac{y_\ell}{|C_\ell|} \sum_{\gamma \in C_\ell} \mu_{ij}(\gamma, 1/2). \quad (16)$$

Now, using part (a) of the proposition,

$$\frac{1}{|C_\ell|} \sum_{\gamma \in C_\ell} \mu_{ij}(\gamma, 1/2) = \frac{1}{|C_\ell|} \sum_{\gamma \in C_\ell \cap \Gamma_{ij}} \mu_{ij}(\gamma, 1/2) = \frac{|C_\ell \cap \Gamma_{ij}|}{|C_\ell|} \alpha_\ell = \frac{2S_\ell \alpha_\ell}{M(M-1)}, \quad (17)$$

where we have made use of Eq. (6) in the last step. We now use Eq. (17) to further simplify Eq. (16) and obtain

$$\Lambda^* = \min_{y_1, \dots, y_L} \sum_{\ell=1}^L \frac{2y_\ell}{M(M-1)} S_\ell \alpha_\ell. \quad (18)$$

Notice that the inequality  $\alpha_\ell \leq 0$  holds for each  $\ell$ . Therefore, an optimal solution to the optimization problem of Eq. (18) is obtained by choosing a class  $C_{\ell^*}$  for which the value of  $S_\ell |\alpha_\ell|$  is maximized and letting  $y_{\ell^*} = 1$ , and  $y_\ell = 0$  if  $\ell \neq \ell^*$ . **Q.E.D.**

Our conclusions are therefore similar to the small noise case. In particular, there exists a best class and all decision rules to be used should belong to a best class. The nature of the best class is interesting. The constant  $\alpha_\ell$  can be interpreted as a measure of the contribution of an "average" element of  $C_\ell$  to a pair of hypotheses which are separated by that decision rule [see Prop. 2(a)]. The product  $S_\ell |\alpha_\ell|$  weighs the number of separations of a decision rule in  $C_\ell$  by the "quality measure"  $\alpha_\ell$  and the value of this product is used to determine a best class.

The identity of the best class cannot be determined analytically because the formulas for the coefficients  $\alpha_\ell$  are somewhat cumbersome. On the other hand, for any given value of  $\epsilon$ , the value of  $\alpha_\ell$  is easy to compute numerically. We have done so for the case where  $D = 2$  and for  $M = 5, 10, 20, 30$  [Poly88]. We summarize the results. When  $\epsilon$  is very small, then the optimal class is the one which partitions evenly the observation set, in agreement with the results of Section 4. Interestingly enough, this same class remains optimal for larger values of  $\epsilon$  as well, up to approximately  $1/M$ . At about that point, the identity of the optimal class changes, and the

optimal class is a most uneven one, namely the class  $C^{1,M-1}$ . This latter class remains the best one for all  $\epsilon$  up to  $1/(M-1)$  (which is the largest allowed value for  $\epsilon$ ).

The case  $\epsilon = 1/(M-1)$  has an interesting interpretation. Here, the probability  $\Pr(y = i | H_j)$  is equal to  $\epsilon$  if  $i \neq j$ , and is zero if  $i = j$ . Thus, an observation  $y = i$  provides absolute proof that  $H_i$  is *not* the true hypotheses. If the sensors use decision rules  $\gamma \in C^{1,M-1}$  of the form  $\gamma(i) = 1$  and  $\gamma(j) = 2$ , for  $j \neq i$ , then a message with the value 1 allows the fusion center to eliminate one of the hypotheses. On the other hand, if decision rules in classes other than  $C^{1,M-1}$  are used, then the fusion center is not able to make unequivocal inferences. This argument suggests that  $C^{1,M-1}$  is the optimal class, as confirmed by our numerical experiments.

## 6. DESIGN OF THE OPTIMAL COMMUNICATION RATE FOR THE SMALL NOISE CASE.

A fundamental design problem in decentralized decision making concerns the choice of the communication rate (or available bandwidth) between the different decision making units. Such design problems are usually very hard and very little analysis is possible, except for simple situations. For this reason, the solution of even idealized problems can provide valuable intuition. We consider such a design problem, in the context of our decentralized detection problem, under the small low noise assumption.

We express the communication rate of each sensor as a function of the variable  $D$ . In particular, we view the number  $\lceil \log_2 D \rceil$  as the number of binary messages that each sensor must transmit to the fusion center<sup>†</sup>. Clearly, a higher value of  $D$  leads to better performance (smaller probability of error at the fusion center) since a decision is made with more information. On the other hand, communication resources may be scarce, in which case an upper bound can be imposed on the total

---

<sup>†</sup> In an alternative formulation we could use  $\log_2 D$  instead of  $\lceil \log_2 D \rceil$ . Which one of these choices is more appropriate could depend on the particular coding method used for transmission. In any case, our subsequent results can be shown to remain valid under this alternative formulation as well.

communication rate in the system. Accordingly, we assume that

$$N \lfloor \log_2 D \rfloor \leq K, \quad (19)$$

where  $K$  is a given positive integer. Given such a constraint, we pose the question: "Is it better to have few sensors communicating at high rate, or more sensors communicating at low rate"? We formulate the above described problem in mathematical terms. We view the optimal error exponent  $\Lambda^*$  as a function of  $D$  and we use the more suggestive notation  $\Lambda^*(D)$ . Furthermore, we consider the small noise case for which we can use the approximation [cf. Eq. (10) and Prop. 1]

$$\Lambda^*(D) = \log \epsilon \frac{S^*(D)}{M(M-1)}, \quad (20)$$

where [cf. Eq. (5)]

$$S^*(D) = \max_{\delta_D \in \Delta_D} \frac{1}{2} \sum_{d=1}^D \delta_d (M - \delta_d), \quad (21)$$

and  $\Delta_D$  is the set of all vectors  $\delta = (\delta_1, \dots, \delta_D)$  such that each  $\delta_d$  is a nonnegative integer and  $\sum_{d=1}^D \delta_d = M$ . Recall that the error probability behaves, asymptotically as  $N \rightarrow \infty$ , like  $e^{N\Lambda^*(D)}$ .

We are then led to the problem

$$\min_D N\Lambda^*(D) \quad (22)$$

subject to the constraint (19). (Of course  $N$  and  $D$  are also constrained to be an integer larger than 1.)

**Proposition 3:** An optimal solution of the problem defined by Eqs. (19) and (22) is given by  $D = 2$ ,  $N = K$ .

**Proof:** We use Eqs. (20) and (21) and the fact that  $\log \epsilon$  is negative to formulate the problem (22) in the form

$$\max_D NF(D), \quad (23)$$

where

$$F(D) = \max_{\delta \in \Delta_D} \sum_{d=1}^D \delta_d (M - \delta_d). \quad (24)$$

Let us recall that the optimization problem in the definition of  $F(D)$  was solved in the end of Section 4. In particular, it is seen that

$$F(2) = \begin{cases} \frac{M^2}{2}, & \text{if } M \text{ is even,} \\ \frac{M^2-1}{2}, & \text{if } M \text{ is odd,} \end{cases}$$

and

$$F(D) \leq \sum_{d=1}^D \frac{M}{D} \left( M - \frac{M}{D} \right) = M^2 \left( 1 - \frac{1}{D} \right), \quad \forall D.$$

[The above inequality is obtained because  $\delta_1 = \dots = \delta_D = M/D$  is the optimal solution in Eq. (24) when the integrality constraints are relaxed.]

We compare the solution  $N = K$ ,  $D = 2$ , with the solution  $N = \lfloor K/2 \rfloor$ ,  $D = 3$ . It is easily verified that

$$K \frac{M^2 - 1}{2} \geq \frac{K}{2} M^2 \left( 1 - \frac{1}{3} \right), \quad \forall M \geq 2,$$

which shows that the solution with  $D = 2$  is preferable. Similarly,

$$K \frac{M^2 - 1}{2} \geq \frac{K}{2} M^2 \left( 1 - \frac{1}{4} \right), \quad \forall M \geq 2,$$

and  $D = 2$  is also preferable to  $D = 4$ . Finally, if  $D > 4$ , then  $\lceil \log_2 D \rceil \geq 3$  and  $N \leq K/3$ . We have

$$K \frac{M^2 - 1}{2} \geq \frac{K}{3} M^2 \geq \frac{K}{3} M^2 \left( 1 - \frac{1}{D} \right), \quad \forall M \geq 2, \forall D > 4,$$

and  $D = 2$  is again preferable. Q.E.D.

Generally speaking, intuition suggests that it is better to have several sensors transmitting low rate but independent information, rather than few sensors transmitting detailed information. The above result corroborates this intuition, at least for the particular problem under study. An alternative statement of this result, which is pertinent to organizations involving human decision makers, is the following: if a decision maker is to receive a set of reports of a given total length, it is preferable to receive many partial but independently drafted reports, rather than a few lengthy ones.

## 7. CONCLUSIONS

We have considered the asymptotic (as the number of sensors goes to infinity) solution of a particularly simple symmetric problem in decentralized detection. While the problem is very idealized, the conclusions obtained agree with intuition and could be useful as guiding principles for more general problems. Roughly stated, the following guidelines suggest themselves:

- a) It is preferable to have several independent sensors transmitting low rate (coarse) information

instead of few sensors transmitting high rate (very detailed) information. (Of course, this guideline is meaningful if it is assumed that the addition of more sensors does not lead to increased "setup" costs; in other words, it is assumed that many sensors are readily available and the only question is whether they can be usefully employed.)

b) An  $M$ -ary hypothesis testing problem can be viewed as a collection of  $M(M - 1)/2$  binary hypothesis testing problem. Under this point of view, the most useful messages by the sensors (decision rules) are those which provide information to the fusion center that is relevant to the largest possible number of binary hypothesis testing problems.

To what extent the above two guidelines can be verified analytically or experimentally in more realistic problems is an interesting question which is left for further research.

## APPENDIX

We outline here a proof that the symmetric solution ( $x_i = 1/3$ , for  $i = 1, 2, 3$ ) is a strict local minimum for the problem considered in the example of Section 5. The problem under consideration can be stated as:

$$\Lambda^* = \min_{x \in X} F(x),$$

where

$$F(x) = \max_{i < j} F_{ij}(x), \quad (\text{A.1})$$

and

$$F_{ij}(x) = \min_{s \in [0,1]} [x_i \nu(s) + x_j \nu(1-s)],$$

where  $i, j \in \{1, 2, 3\}$ . Let  $x^* = (1/3, 1/3, 1/3)$ . The function  $\nu(\cdot)$  is strictly convex and continuously differentiable, and the minimum in the definition of  $F_{ij}(x^*)$  is uniquely attained at  $s = 1/2$ . We can then use Danskin's Theorem [Dan67] to obtain

$$\frac{\partial F_{ij}}{\partial x_k}(x^*) = \frac{\partial}{\partial x_k} \left[ x_i \nu\left(\frac{1}{2}\right) + x_j \nu\left(\frac{1}{2}\right) \right] \Big|_{x=x^*} = \begin{cases} 0, & \text{if } k \neq i \text{ and } k \neq j; \\ \nu\left(\frac{1}{2}\right), & \text{if } k = i \text{ or } k = j. \end{cases}$$

Consider any direction  $d \in R^3, d \neq 0$ , in which  $x^*$  can be perturbed without leaving the set  $X$ . [That is,  $d = (d_1, d_2, d_3)$  with  $d_1 + d_2 + d_3 = 0$ .] The chain rule yields

$$\frac{\partial F_{ij}(x^* + \alpha d)}{\partial \alpha} \Big|_{\alpha=0} = \sum_{k=1}^3 d_k \frac{\partial F_{ij}(x^*)}{\partial x_k} = (d_i + d_j) \nu\left(\frac{1}{2}\right). \quad (\text{A.2})$$

Notice that the assumptions  $d \neq 0$ ,  $d_1 + d_2 + d_3 = 0$  imply that there exist some  $i, j$  such that  $d_i + d_j < 0$ . Since  $\nu(1/2) < 0$ , it follows that for every choice of  $d$ , the left-hand side of Eq. (A.2) is positive for some pair  $(i, j)$ . Thus, for each direction  $d$ , some function  $F_{i,j}(x)$  has to increase. Taking Eq. (A.1) into account,  $F(x)$  must also increase. From this point on, it is only a small step to show that  $F(x)$  is larger than  $F^*(x)$  in a neighborhood of  $x^*$ , i.e., that  $x^*$  is a local minimum. (The details of this last step are omitted.)

## REFERENCES

- [ChV86] Chair, Z., and P.K. Varshney, "Optimum Data Fusion in Multiple Sensor Detection Systems" *IEEE Transactions on Aerospace and Electronic Systems*, AES-22, 1986, pp. 98-101.
- [Dan67] Danskin, J.M., *Theory of Max-Min and its Application in Weapons Allocation Problems*, Springer-Verlag, 1967.
- [EkT82] Ekchian, L.K., and R.R. Tenney, "Detection Networks", *Proceedings of the 21st IEEE Conference on Decision and Control*, 1982, pp. 686-691.
- [HoV86] Hoballah, I.Y., and P.K. Varshney, "Neyman-Pearson Detection with Distributed Sensors", *Proceedings of the 25th IEEE Conference on Decision and Control*, Athens, Greece, December 1986, pp. 237-241.
- [KuP82] Kushner, H.J., and A. Pacut, "A Simulation Study of a Decentralized Detection Problem", *IEEE Transactions on Automatic Control*, 27, 1982, pp. 1116-1119.
- [PaA86] Papastavrou, J.D., and M. Athans, "A Distributed Hypothesis-Testing Team Decision Problem with Communications Cost", *Proceedings of the 25th IEEE Conference on Decision and Control*, Athens, Greece, December 1986, pp. 219-225.
- [Pol88] Polychronopoulos, G. H., "Solution of Some Problems in Decentralized Detection by a Large Number of Sensors", Technical Report LIDS-TH-1781, Laboratory for Information and Decision Systems, M.I.T., Cambridge, MA, June 1988.
- [ReN87a] Reibman, A.R., and L.W. Nolte, "Optimal Detection and Performance of Distributed Sensor Systems", *IEEE Transactions on Aerospace and Electronic Systems*, AES-23, 1987, pp. 24-30.
- [ReN87b] Reibman, A.R., and L.W. Nolte, "Design and Performance Comparison of Distributed Detection



Networks", *IEEE Transactions on Aerospace and Electronic Systems*, AES-23, 1987, pp. 789-797.

- [Sad86] Sadjadi, F., "Hypothesis Testing in a Distributed Environment", *IEEE Transactions on Aerospace and Electronic Systems*, AES-22, 1986, pp. 134-137.
- [SGB67] Shannon, C.E., Gallager, R.G., and Berlekamp E.R., "Lower Bounds to Error Probability for Coding on Discrete Memoryless Channels", *Information and Control*, 10, 1967, pp. 65 -103.
- [Sri86a] Srinivasan, R., "Distributed Radar Detection Theory", *IEE Proceedings*, 133, 1986, pp. 55-60.
- [Sri86b] Srinivasan, R., "A Theory of Distributed Detection", *Signal Processing*, 11, 1986, pp. 319-327.
- [TeS81] Tenney, R.R., Sandell, N.R. Jr., "Detection with Distributed Sensors", *IEEE Transactions on Aerospace and Electronic Systems*, AES-17, 4, 1981, pp. 501-510.
- [TeV84] Teneketzis, D., and Varaiya, P., "The Decentralized Quickest Detection Problem", *IEEE Transactions on Automatic Control*, AC-29, 1984, pp. 641-644.
- [TVB87] Thomopoulos, S.C.A., R. Viswanathan, and D.P. Bougoulas, "Optimal Decision Fusion in Multiple Sensor Systems", *IEEE Transactions on Aerospace and Electronic Systems*, AES-23, 1987.
- [TsA85] Tsitsiklis, J.N., Athans M., "On the Complexity of Decentralized Decision Making and Detection Problems", *IEEE Transactions on Automatic Control*, AC-30, 5, 1985, pp. 440-446.
- [Tsi84] Tsitsiklis, J.N., "Problems in Decentralized Decision Making and Computation", Ph.D. Thesis, Dept. of Electrical Engineering and Computer Science, M.I.T., Cambridge, Massachusetts, 1984.
- [Tsi 88] Tsitsiklis, J.N., "Decentralized Detection by a Large Number of Sensors", *Mathematics of Control, Signals, and Systems*, 1,2, 1988, pp. ??-??.
- [VaT68] Van Trees, H. L., *Detection Estimation and Modulation Theory*, Vol. I, J. Wiley & Sons, New York, 1968.